

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**



501.43636X00

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicants: Satoshi YAGI

Serial No.: 10/795,996

Filing Date: March 10, 2004

For: CONTROL METHOD FOR DATA TRANSFER DEVICE, DATA TRANSFER CIRCUIT, AND DISK ARRAY DEVICE

LETTER CLAIMING RIGHT OF PRIORITY

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

June 3, 2004

Sir:

Under the provisions of 35 USC 119 and 37 CFR 1.55, applicants hereby claim the right of priority based on:

**Japanese Application No. 2003-314010
Filed: September 5, 2003**

A Certified copy of said application document is attached hereto.

Acknowledgement thereof is respectfully requested.

Respectfully submitted,


Carl I. Brundidge
Registration No. 29,621
ANTONELLI, TERRY, STOUT & KRAUS, LLP

CIB/jdc
Enclosures
703/312-6600

日本国特許庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日 2003年 9月 5日
Date of Application:

出願番号 特願2003-314010
Application Number:

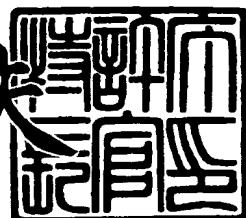
[ST. 10/C] : [JP2003-314010]

出願人 株式会社日立製作所
Applicant(s):

2004年 3月 2日

特許庁長官
Commissioner,
Japan Patent Office

今井康夫



出証番号 出証特2004-3015597

●

【書類名】 特許願
【整理番号】 HI030532
【提出日】 平成15年 9月 5日
【あて先】 特許庁長官殿
【国際特許分類】 G06F 3/06
【発明者】
【住所又は居所】 神奈川県小田原市中里 322番2号 株式会社日立製作所 R A I
D システム事業部内
【氏名】 八木 聰
【特許出願人】
【識別番号】 000005108
【氏名又は名称】 株式会社日立製作所
【代理人】
【識別番号】 110000176
【氏名又は名称】 一色国際特許業務法人
【代表者】 一色 健輔
【手数料の表示】
【予納台帳番号】 211868
【納付金額】 21,000円
【提出物件の目録】
【物件名】 特許請求の範囲 1
【物件名】 明細書 1
【物件名】 図面 1
【物件名】 要約書 1

【書類名】特許請求の範囲**【請求項1】**

記憶装置に対する書き込みデータを受信するデータ受信部と、

前記データ受信部が受信する前記書き込みデータを前記記憶装置に転送するデータ制御部と、

前記記憶装置の記憶領域上に記憶される一連のデータを記憶するデータ記憶部と、
を備えるデータ転送装置の制御方法であって、

前記データ制御部は、前記記憶装置に記憶されている前記一連のデータをブロック単位で読み出して前記データ記憶部に記憶し、

前記データ制御部は、受信した前記書き込みデータについて、当該書き込みデータの書き込み先となる前記記憶装置の記憶領域上のブロックと、前記データ記憶部に記憶されている前記記憶装置の記憶領域上の前記ブロックとが同じ場合は、前記書き込みデータの書き込み先となる前記記憶装置の記憶位置に対応する前記データ記憶部に記憶されているデータを前記書き込みデータにより更新し、

前記データ制御部は、受信した前記書き込みデータについて、当該書き込みデータの書き込み先となる前記記憶装置の記憶領域上のブロックと、前記データ記憶部に記憶されている前記記憶装置の記憶領域上の前記ブロックとが異なる場合は、前記データ記憶部に記憶されている一連のデータに基づいて保証コードを生成し、前記データ記憶部に記憶されている一連のデータに前記生成した前記保証コードを付加して前記記憶装置に転送し、前記書き込みデータの書き込み先となる前記記憶装置の記憶領域上のブロックに記憶されている一連のデータを読み出して前記データ記憶部に記憶し、前記書き込みデータの書き込み先となる前記記憶装置の記憶位置に対応する前記データ記憶部に記憶されているデータを前記書き込みデータにより更新すること、

を特徴とするデータ転送装置の制御方法。

【請求項2】

前記データ受信部により連続して受信される一連の前記書き込みデータは、前記ブロックの連続する領域に前記受信される順序では必ずしも書き込まれないこと、

を特徴とする請求項1に記載のデータ転送装置の制御方法。

【請求項3】

記憶装置に対する書き込みデータを受信するデータ受信部と、

前記データ受信部が受信する前記書き込みデータを前記記憶装置に転送するデータ制御部と、

前記記憶装置の記憶領域上に記憶される一連のデータを記憶するデータ記憶部と、
を備え、

前記データ制御部は、前記記憶装置に記憶されている前記一連のデータをブロック単位で読み出して前記データ記憶部に記憶し、

前記データ制御部は、受信した前記書き込みデータについて、当該書き込みデータの書き込み先となる前記記憶装置の記憶領域上のブロックと、前記データ記憶部に記憶されている前記記憶装置の記憶領域上の前記ブロックとが同じ場合は、前記書き込みデータの書き込み先となる前記記憶装置の記憶位置に対応する前記データ記憶部に記憶されているデータを前記書き込みデータにより更新し、

前記データ制御部は、受信した前記書き込みデータについて、当該書き込みデータの書き込み先となる前記記憶装置の記憶領域上のブロックと、前記データ記憶部に記憶されている前記記憶装置の記憶領域上の前記ブロックとが異なる場合は、前記データ記憶部に記憶されている一連のデータに基づいて保証コードを生成し、前記データ記憶部に記憶されている一連のデータに前記生成した前記保証コードを付加して前記記憶装置に転送し、前記書き込みデータの書き込み先となる前記記憶装置の記憶領域上のブロックに記憶されている一連のデータを読み出して前記データ記憶部に記憶し、前記書き込みデータの書き込み先となる前記記憶装置の記憶位置に対応する前記データ記憶部に記憶されているデータを前記書き込みデータにより更新すること、

を特徴とするデータ転送回路。

【請求項4】

前記データ受信部により連続して受信される一連の前記書き込みデータは、前記ブロックの連続する領域に前記受信される順序では必ずしも書き込まれないこと、

を特徴とする請求項3に記載のデータ転送回路。

【請求項5】

情報処理装置からディスクドライブに対する書き込みデータを受信するホストインタフェースと、

前記ホストインタフェースが受信する前記書き込みデータを前記ディスクドライブに転送するデータコントローラと、を備えるディスクアレイ装置であって、

前記データコントローラは、前記ディスクドライブに対する書き込みデータを前記ホストインタフェースから受信するデータ受信部と、前記データ受信部が受信する前記書き込みデータを前記ディスクドライブに転送するデータ制御部と、前記ディスクドライブに記憶される一連のデータを記憶するデータ記憶部と、を備え、

前記データ制御部は、前記ディスクドライブに記憶されている前記一連のデータをブロック単位で読み出して前記データ記憶部に記憶し、

前記データ制御部は、受信した前記書き込みデータについて、当該書き込みデータの書き込み先となる前記ディスクドライブの記憶領域上のブロックと、前記データ記憶部に記憶されている前記ディスクドライブの記憶領域上の前記ブロックとが同じ場合は、前記書き込みデータの書き込み先となる前記ディスクドライブの記憶位置に対応する前記データ記憶部に記憶されているデータを前記書き込みデータにより更新し、

前記データ制御部は、受信した前記書き込みデータについて、当該書き込みデータの書き込み先となる前記ディスクドライブの記憶領域上のブロックと、前記データ記憶部に記憶されている前記ディスクドライブの記憶領域上の前記ブロックとが異なる場合は、前記データ記憶部に記憶されている一連のデータに基づいて保証コードを生成し、前記データ記憶部に記憶されている一連のデータに前記生成した前記保証コードを付加して前記ディスクドライブに転送し、前記書き込みデータの書き込み先となる前記ディスクドライブの記憶領域上のブロックに記憶されている一連のデータを読み出して前記データ記憶部に記憶し、前記書き込みデータの書き込み先となる前記ディスクドライブの記憶位置に対応する前記データ記憶部に記憶されているデータを前記書き込みデータにより更新すること、

を特徴とするディスクアレイ装置。

【請求項6】

前記ホストインタフェースは、ネットワークを介して前記情報処理装置と接続し、

前記ホストインタフェースにより連続して受信される一連の前記書き込みデータは、前記ブロックの連続する領域に前記受信される順序では必ずしも書き込まれないこと、

を特徴とする請求項5に記載のディスクアレイ装置。

【請求項7】

情報処理装置からディスクドライブに対する書き込みデータを受信するホストインタフェースと、

前記ホストインタフェースが受信する前記書き込みデータを前記ディスクドライブに転送するデータコントローラと、

全体の制御を制御部を司るプロセッサと、

データを記憶するメモリと、を備えるディスクアレイ装置であって、

前記プロセッサは、前記ディスクドライブに記憶されている前記一連のデータをブロック単位で読み出して前記メモリに記憶し、

前記プロセッサは、受信した前記書き込みデータについて、当該書き込みデータの書き込み先となる前記ディスクドライブの記憶領域上のブロックと、前記メモリに記憶されている前記ディスクドライブの記憶領域上の前記ブロックとが同じ場合は、前記書き込みデータの書き込み先となる前記ディスクドライブの記憶位置に対応する前記メモリに記憶されているデータを前記書き込みデータにより更新し、

前記プロセッサは、受信した前記書き込みデータについて、当該書き込みデータの書き込み先となる前記ディスクドライブの記憶領域上のブロックと、前記メモリに記憶されている前記ディスクドライブの記憶領域上の前記ブロックとが異なる場合は、前記メモリに記憶されている一連のデータに基づいて保証コードを生成し、前記メモリに記憶されている一連のデータに前記生成した前記保証コードを附加して前記ディスクドライブに転送し、前記書き込みデータの書き込み先となる前記ディスクドライブの記憶領域上のブロックに記憶されている一連のデータを読み出して前記メモリに記憶し、前記書き込みデータの書き込み先となる前記ディスクドライブの記憶位置に対応する前記メモリに記憶されているデータを前記書き込みデータにより更新すること、

を特徴とするディスクアレイ装置。

【請求項8】

情報処理装置からディスクドライブに対する書き込みデータを受信するチャネル制御部と、

前記ディスクドライブに対するデータの書き込みに関する処理を行うディスク制御部と、

前記チャネル制御部と前記ディスク制御部との間で授受されるデータを記憶するキャッシュメモリと、を備えるディスクアレイ装置であって、

前記チャネル制御部は、前記書き込みデータを受信するデータ受信部と、前記データ受信部が受信する前記書き込みデータを前記キャッシュメモリに転送するデータ制御部と、前記ディスクドライブの記憶領域上に記憶される一連のデータを記憶するデータ記憶部とを備え、

前記データ制御部は、前記ディスクドライブに記憶されている前記一連のデータを前記キャッシュメモリからブロック単位で読み出して前記データ記憶部に記憶し、

前記データ制御部は、受信した前記書き込みデータについて、当該書き込みデータの書き込み先となる前記ディスクドライブの記憶領域上のブロックと、前記データ記憶部に記憶されている前記ディスクドライブの記憶領域上の前記ブロックとが同じ場合は、前記書き込みデータの書き込み先となる前記ディスクドライブの記憶位置に対応する前記データ記憶部に記憶されているデータを前記書き込みデータにより更新し、

前記データ制御部は、受信した前記書き込みデータについて、当該書き込みデータの書き込み先となる前記ディスクドライブの記憶領域上のブロックと、前記データ記憶部に記憶されている前記ディスクドライブの記憶領域上の前記ブロックとが異なる場合は、前記データ記憶部に記憶されている一連のデータに基づいて保証コードを生成し、前記データ記憶部に記憶されている一連のデータに前記生成した前記保証コードを附加して前記キャッシュメモリに転送し、前記書き込みデータの書き込み先となる前記ディスクドライブの記憶領域上のブロックに記憶されている一連のデータをキャッシュメモリから読み出して前記データ記憶部に記憶し、前記書き込みデータの書き込み先となる前記ディスクドライブの記憶位置に対応する前記データ記憶部に記憶されているデータを前記書き込みデータにより更新すること、

を特徴とするディスクアレイ装置。

【請求項9】

前記チャネル制御部は、ネットワークを介して前記情報処理装置と接続し、前記書き込みデータを受信するインタフェースを備え、

前記データ受信部は、前記インタフェースから前記書き込みデータを受信し、

前記インタフェースにより連続して受信される一連の前記書き込みデータは、前記ブロックの連続する領域に前記受信される順序では必ずしも書き込まれないこと、

を特徴とする請求項8に記載のディスクアレイ装置。

【請求項10】

情報処理装置からディスクドライブに対する書き込みデータを受信するチャネル制御部と、

前記ディスクドライブに対するデータの書き込みに関する処理を行うディスク制御部と、

前記チャネル制御部と前記ディスク制御部との間で授受されるデータを記憶するキャッシュメモリと、を備えるディスクアレイ装置であって、

前記ディスク制御部は、前記書き込みデータをキャッシュメモリから読み出すデータ読み出し部と、前記データ読み出し部が読み出す前記書き込みデータを前記ディスクドライブに転送するデータ制御部と、前記ディスクドライブの記憶領域上に記憶される一連のデータを記憶するデータ記憶部とを備え、

前記データ制御部は、前記ディスクドライブに記憶されている前記一連のデータをブロック単位で読み出して前記データ記憶部に記憶し、

前記データ制御部は、前記キャッシュメモリから読み出した前記書き込みデータについて、当該書き込みデータの書き込み先となる前記ディスクドライブの記憶領域上のブロックと、前記データ記憶部に記憶されている前記ディスクドライブの記憶領域上の前記ブロックとが同じ場合は、前記書き込みデータの書き込み先となる前記ディスクドライブの記憶位置に対応する前記データ記憶部に記憶されているデータを前記書き込みデータにより更新し、

前記データ制御部は、前記キャッシュメモリから読み出した前記書き込みデータについて、当該書き込みデータの書き込み先となる前記ディスクドライブの記憶領域上のブロックと、前記データ記憶部に記憶されている前記ディスクドライブの記憶領域上の前記ブロックとが異なる場合は、前記データ記憶部に記憶されている一連のデータに基づいて保証コードを生成し、前記データ記憶部に記憶されている一連のデータに前記生成した前記保証コードを附加して前記ディスクドライブに転送し、前記書き込みデータの書き込み先となる前記ディスクドライブの記憶領域上のブロックに記憶されている一連のデータを読み出して前記データ記憶部に記憶し、前記書き込みデータの書き込み先となる前記ディスクドライブの記憶位置に対応する前記データ記憶部に記憶されているデータを前記書き込みデータにより更新すること、

を特徴とするディスクアレイ装置。

【書類名】明細書

【発明の名称】データ転送装置の制御方法、データ転送回路、及びディスクアレイ装置

【技術分野】

【0001】

この発明は、データ転送装置の制御方法、データ転送回路、及びディスクアレイ装置に関する。

【背景技術】

【0002】

近年の情報処理システムが取り扱うデータの内容はますます重要度を増し、記憶装置に対して信頼性が求められている。記憶装置の信頼性を高める方法として、ハードディスク等の記憶デバイスに記憶されるデータにLRC (Longitudinal Redundancy Check) 符号などの保証コードを附加する方法が知られている。記憶装置と情報処理装置との間でのデータのやり取りはブロック単位で行われ、保証コードもブロック単位に附加されている。

【特許文献1】特許3186272号公報

【発明の開示】

【発明が解決しようとする課題】

【0003】

ところで、近年では記憶装置に接続する方式が多様化し、ファイバーチャネルやiSCSI (Internet SCSI) のように、ネットワークによって情報処理装置と記憶装置とをつなぐ技術が採用され始めている。ネットワークによって情報処理装置と記憶装置とが接続される環境において、情報処理装置と記憶装置とが複数の経路で接続されたり、情報処理装置から記憶装置に送信されるデータのブロックがネットワーク上ではそのブロックの大きさとは異なるサイズでデータが伝送されたりすることがある。そのため、情報処理装置から送信されるブロックの一部だけが記憶装置に到達するような状況が起こりうる。また、あるブロックの一部と、そのブロックとは異なるブロックの一部とが混在して伝送されることもありうる。このように、記憶装置がデータのブロックを、記憶装置の連続した領域に書き込む順序に一連のデータとして受信できないような状況が起こりうる。

【0004】

しかしながら、従来の記憶装置においては、少なくとも1ブロック分のデータが、記憶装置に記憶されるアドレス順に連続して伝送されることが前提とされており、上記のネットワーク接続の形態のように、ブロックの一部だけが伝達されるような状況への対応ができなかった。

【0005】

本発明はこのような状況を鑑みてなされたものであり、記憶装置が記憶するデータのブロックについての保証コードをブロック単位に計算する、データ転送装置の制御方法、データ転送回路、及びディスクアレイ装置を提供することを目的とする。

【課題を解決するための手段】

【0006】

上記目的を達成する本発明のうち主たる発明は、記憶装置に対する書き込みデータを受信するデータ受信部と、前記データ受信部が受信する前記書き込みデータを前記記憶装置に転送するデータ制御部と、前記記憶装置の記憶領域上に記憶される一連のデータを記憶するデータ記憶部と、を備えるデータ転送装置の制御方法であって、前記データ制御部は、前記記憶装置に記憶されている前記一連のデータをブロック単位で読み出して前記データ記憶部に記憶し、前記データ制御部は、受信した前記書き込みデータについて、当該書き込みデータの書き込み先となる前記記憶装置の記憶領域上のブロックと、前記データ記憶部に記憶されている前記記憶装置の記憶領域上の前記ブロックと同じ場合は、前記書き込みデータの書き込み先となる前記記憶装置の記憶位置に対応する前記データ記憶部に記憶されているデータを前記書き込みデータにより更新し、前記データ制御部は、受信した前記書き込みデータについて、当該書き込みデータの書き込み先となる前記記憶装置の記憶領域上のブロックと、前記データ記憶部に記憶されている前記記憶装置の記憶領域上

の前記ブロックとが異なる場合は、前記データ記憶部に記憶されている一連のデータに基づいて保証コードを生成し、前記データ記憶部に記憶されている一連のデータに前記生成した前記保証コードを付加して前記記憶装置に転送し、前記書き込みデータの書き込み先となる前記記憶装置の記憶領域上のブロックに記憶されている一連のデータを読み出して前記データ記憶部に記憶し、前記書き込みデータの書き込み先となる前記記憶装置の記憶位置に対応する前記データ記憶部に記憶されているデータを前記書き込みデータにより更新すること、を特徴とする。

【0007】

ここで、前記データ転送装置とは、例えば後述するディスクアレイ装置200の備えるデータコントローラ500である。前記記憶装置とは、例えば後述するディスクアレイ装置200の備えるキャッシュメモリ206やディスクドライブ300である。前記保証コードとは、例えば、LRC符号やVRC符号、CRC符号、パリティ、チェックサム、ハミング符号などの誤り検出符合である。前記記憶装置に記憶されるデータはブロック単位に読み出され、ブロックを形成する一連のデータに対して保証コードの計算が行われ、計算された保証コードと前記ブロック単位に付加された保証コードとを用いて、ブロック単位のデータに誤りがないかどうかが検査される。したがって、前記記憶装置にはブロック単位にデータが記憶され、各ブロックを形成する一連のデータに対して、読み出されるときと同様の計算によって求められる保証コードが付加されている必要がある。

【0008】

本発明によれば、データ制御部は記憶装置に記憶されている一連のデータをデータ記憶部に読み出し、読み出した一連のデータを受信した書き込みデータにより更新し、更新した一連のデータに基づいて保証コードを計算する。したがって、データ制御部はデータ記憶部に記憶されている一連のデータについて、記憶装置に記憶されている記憶位置の順番どおりに保証コードの計算を行うことができる。よって、記憶装置に記憶される記憶位置の順番とは異なる順番でデータ受信部が複数の書き込みデータを受信した場合にも、データ制御部はデータ記憶部に記憶されている一連のデータに基づき、記憶装置に記憶されている記憶位置の順番どおりに保証コードを計算することができる。また、データ受信部が受信する書き込みデータがブロックの一部のみを形成する可変長のデータであっても、データ記憶部にはブロック単位でのデータが記憶されているため、データ制御部はブロック単位に保証コードを計算することができる。データ制御部は、書き込み対象となるブロックの一連のデータを適宜記憶装置から読み出すため、データ受信部が同一のブロックに対する書き込みデータを連続して受信できない場合であっても、データ制御部はブロック単位に保証コードを付加することができる。データ制御部はデータ記憶装置に記憶されている一連のデータに基づいて計算した保証コードを、この一連のデータに付加して記憶装置に転送するため、記憶装置には常に記憶位置の順番にしたがって計算された保証コードが付加されたブロック単位のデータが記憶されることになる。

【0009】

なお、本発明の他の特徴については、本明細書及び添付図面の記載により明らかにされる。

【発明の効果】

【0010】

本発明によれば、データ転送装置の制御方法、データ転送回路、及びディスクアレイ装置を提供することができる。

【発明を実施するための最良の形態】

【0011】

以下に本発明の実施形態について図面を用いて詳細に説明する。

【0012】

====第1の実施の形態====

図1は本実施の形態にかかるデータ転送回路を備えるディスクアレイ装置を含む情報処理システムの構成を示すブロック図である。データ転送回路とは後述するデータコントロ

ーラである。図1に示す情報処理システムは、各種の情報処理サービスを提供する情報処理装置100と、ディスクドライブ300の記憶領域を情報処理装置100に提供するディスクアレイ装置200とを含んで構成される。

【0013】

情報処理装置100はCPU (Central Processing Unit) やメモリを備えたコンピュータである。情報処理装置100では、情報処理装置100の備えるCPUによって各種のプログラムが実行されることにより様々な機能が実現される。情報処理装置100は例えばパーソナルコンピュータやワークステーション、メインフレームコンピュータなどである。情報処理装置100は1台のコンピュータであってもよいし、複数台のコンピュータであってもよい。情報処理装置100上ではオペレーティングシステムが実行され、オペレーティングシステム上では様々なアプリケーションプログラムが実行される。

【0014】

情報処理装置100はSAN (Storage Area Network) 400を介してディスクアレイ装置200と接続されている。SAN 400を介して行われる情報処理装置100とディスクアレイ装置200との間の通信は、一般的にファイバチャネルプロトコルに従って行われるが、SAN 400はファイバチャネルプロトコル以外にも様々なプロトコルに従った通信経路とすることができる。情報処理装置100とディスクアレイ装置200とを接続する通信形態には、例えば、LAN (Local Area Network) やSCSI (Small Computer System Interface) 、iSCSI (Internet Small Computer System Interface) 、ESCON (Enterprise System Connection) (登録商標) 、FICON (Fibre Connection) (登録商標) 、ACONARC (Advanced Connection Architecture) (登録商標) 、FIBARC (Fibre connection Architecture) (登録商標) などを用いることができる。また、情報処理装置100とディスクアレイ装置200とはネットワーク接続の形態ではなく、直接接続する形態とすることもできる。

【0015】

情報処理装置100はファイバチャネルプロトコルに従って、ブロック単位でのデータ入出力の要求（以下、ブロックアクセス要求と称する）をディスクアレイ装置200に送信する。ディスクアレイ装置200は、情報処理装置100からブロックアクセス要求を受信すると、受信したブロックアクセス要求に応じて、ディスクドライブ300に対するデータの入出力に関する処理を行う。このようにして、情報処理装置100上で実行されるアプリケーションプログラムは、ディスクドライブ300の記憶領域に適宜アクセスしながら、各種処理を実行する。

【0016】

ディスクアレイ装置200は多数の物理ディスク（ディスクドライブ300）を備え、複数のディスクドライブ300による記憶領域が情報処理装置100に提供されている。ディスクドライブ300としては例えば、ハードディスク装置やフレキシブルディスク装置、半導体記憶装置等を用いることができる。ディスクドライブ300により提供される物理的な記憶領域は、論理的に設定される論理ユニット（Logical Unit）を単位として管理されている。各論理ユニットには論理ユニット番号（Logical Unit Number;LUN）という識別子が付与され、ディスクドライブ300に対するデータ入出力は、論理ユニット番号を指定して行われる。ディスクアレイ装置200は、ディスクドライブ300に設定される論理ユニット上に、例えば512バイトといったブロック単位でデータを管理する。ディスクアレイ装置200はこのブロック単位でデータの入出力に関する処理を行う。前記の単位となるブロックのことは論理ブロックと呼ばれている。各論理ブロックには記憶位置情報を示す論理ブロックアドレス（Logical Block Address;LBA）が付与され、情報処理装置100からディスクアレイ装置200に対するブロックアクセス要求には、論理ユニット番号と論理ブロックアドレスとが指定される。

【0017】

ディスクアレイ装置200は、複数のディスクドライブ300によってディスクアレイを構成し、RAIDによって管理される記憶領域を提供することもできるし、単一の物理

ディスクによる記憶領域を提供するようにもできる。ディスクドライブ300はディスクアレイ装置200に一体的に構成されてもよいし、ディスクアレイ装置200から独立した装置として、SCSIやLAN、SANといった通信経路によってディスクアレイ装置200と接続する形態としてもよい。

【0018】

====ディスクアレイ装置====

ディスクアレイ装置200はCPU201、メモリ202、NVRAM203、FPC204、FPC205、キャッシュメモリ206、データコントローラ500を備えている。これらは互いにバス207を介して接続されている。バス207は、例えば、PCIバスであり、32ビット単位でデータを伝送する。バス207で伝送されるデータの単位はワードと呼ばれる。もちろんバス207はPCIバス以外のバスであってもよい。また1ワードは32ビットに限らず、8ビットや16ビット、64ビット、128ビットなどの単位とすることができる。

【0019】

CPU201は、ディスクアレイ装置200の全体の制御を司るもので、メモリ202やNVRAM203に格納されたアプリケーションプログラムを実行することにより、ディスクドライブ300の管理やロックアクセス要求の解釈など様々な機能を実現することができる。NVRAM203は、CPU201が実行するプログラムを格納する、例えばフラッシュメモリなどの不揮発性メモリである。

【0020】

FPC204は情報処理装置100との間で通信を行うインターフェースである。FPC204はホストインターフェースとも呼ばれる。FPC204は、ファイバチャネルプロトコルに従ってロックアクセス要求を受け付ける機能を備える。FPC204はSAN400の通信形態によって、例えば、iSCSIなどのプロトコルによるロックアクセス要求を受け付けるようにすることもできる。

【0021】

FPC205はディスクドライブ300との間でデータのやり取りを行うインターフェースである。FPC205はディスクインターフェースとも呼ばれる。FPC205はディスクドライブ300を制御するコマンドなどを規定するプロトコルに従ってディスクドライブ300に対するデータ入出力要求を送信する機能を備える。FPC205は、例えば、SCSIやファイバチャネルなどのプロトコルに従って、ディスクドライブ300に対してデータの書き込みや読み出しのコマンドを送信することができる。

【0022】

キャッシュメモリ206は、FPC204とFPC205との間で授受されるデータを記憶するためのメモリである。

【0023】

データコントローラ500はCPU201の制御によりFPC204とキャッシュメモリ206との間、あるいはキャッシュメモリ206とFPC205との間でのデータの転送を行うものである。データコントローラ500は、例えば、特定用途向けICにロジックを形成する回路とすることができます。

【0024】

情報処理装置100がディスクドライブ300に対する書き込みを要求するロックアクセス要求（以下、ロック書き込み要求と称する）をディスクアレイ装置200に送信すると、ディスクアレイ装置200では、FPC204がロック書き込み要求を受け付け、データコントローラ500がロック書き込み要求に付随するロック単位の書き込みデータをキャッシュメモリ206に転送する。書き込みデータがキャッシュメモリ206に転送されると、データコントローラ500がキャッシュメモリ206からFPC205に書き込みデータを読み出し、FPC205がディスクドライブ300に対するロック単位での書き込みを指示するコマンドを送信する。

【0025】

====データコントローラ====

図2はディスクアレイ装置200の備えるデータコントローラ500のブロック図である。データコントローラ500は、バスインターフェース部501、キャッシングインターフェース部502、データ制御部503、データ記憶部504を備えている。

【0026】

バスインターフェース部501は、バス207に接続するインターフェースである。バスインターフェース部501はバス207を介してFPC204やFPC205との間でデータの授受を行うデータ受信部及びデータ送信部として機能する。バスインターフェース部501は、FPC204あるいはFPC205からデータを受信するときは、受信したデータに付加されたパリティを用いて受信データの完全性を検査する。またバスインターフェース部501はFPC204あるいはFPC205に対してデータを送信するときには、送信データにパリティを付加して送信する。

【0027】

キャッシングインターフェース部502は、キャッシングメモリ206との間でデータの授受を行うインターフェースである。キャッシングインターフェース部502は、キャッシングメモリ206からデータを読み出すデータ読み出し部、及びキャッシングメモリ206にデータを書き込むデータ書き込み部として機能する。キャッシングインターフェース部502は、キャッシングメモリ206に対してデータを書き込む際にはECC (Error Correcting Code) を計算して付加し、キャッシングメモリ206からデータを読み出す際にはECCによってデータの完全性を検査する。

【0028】

データ制御部503は、FPC204又はFPC205と、キャッシングメモリ206との間でデータを転送機能を有している。

【0029】

データ記憶部504は、1ブロック分の大きさのデータを記憶することができる。もちろん、データ記憶部504に複数のブロック分のデータを記憶するようにしてもよい。データ制御部503はディスクドライブ300からデータを1ブロック読み出し、データ記憶部504に記憶する。データ制御部503は、データ記憶部504に記憶されたデータのブロックについて、32ビット毎のデータに対して排他的論理和演算を用いて保証コードを計算する。

【0030】

====保証コードが付加されたデータ転送処理====

ディスクアレイ装置200では、ディスクドライブ300に記憶されているデータの保全性を保証するために、ディスクドライブ300に対してデータを読み書きする単位である論理ブロック毎に保証コードを付加して管理する。論理ブロックへの保証コードの付加はデータコントローラ500によって行われる。また、論理ブロックに保証コードを付加する以外にも、書き込みデータがディスクドライブ300に書き込まれるまでのデータ転送処理において保証コードが付与されている。以下に、ディスクアレイ装置200内で行われるデータ転送処理について説明する。

【0031】

図3は、情報処理装置100からブロック書き込み要求を受信し、ディスクドライブ300に書き込むまでの流れを示す説明図である。情報処理装置100からは1ブロックの書き込みデータを付随させたブロック書き込み要求が送信されている。

【0032】

FPC204は情報処理装置からのブロック書き込み要求を受信する。情報処理装置100が送信するブロックアクセス要求にはCRC (Cyclic Redundancy Check) 符号が付加されている。FPC204はブロック書き込み要求に付随する書き込みデータについてCRC符号を計算する。FPC204は計算したCRC符号と、ブロックアクセス要求に付加されたCRC符号とを比較することによって、書き込みデータに誤りがないことを確認する (S3001)。FPC204は書き込みデータについてパリティを計算し、書き込み

データにパリティを付加してデータコントローラ500に送信する(S3002)。データコントローラ500は書き込みデータを受信すると、書き込みデータに付加されているパリティを用いて受信したデータが正しいことを確認する(S3003)。データコントローラ500は書き込みデータに対するLRC(Longitudinal Redundancy Check)コードを計算し、書き込みデータに付加する(S3004)。データコントローラ500はLRCコードが付加された書き込みデータについてECCを生成し(S3005)、LRCコードが付加された書き込みデータとともにキャッシュメモリ206に転送する。ここまでの一連の処理で、ディスクアレイ装置200が情報処理装置100から受信したディスクドライブ300に対する書き込みデータは、LRCコードが付加されてキャッシュメモリ206に記憶される。ディスクアレイ装置200はキャッシュメモリ206に書き込みデータを記憶すると、情報処理装置100には書き込みが完了した旨を通知する。

【0033】

一方ディスクアレイ装置200では、キャッシュメモリ206に記憶されているディスクドライブ300に対する書き込みデータは、所定のタイミングでキャッシュメモリ206からディスクドライブ300に転送される。所定のタイミングとは、例えば、ある一定の時間が経過した場合や、キャッシュメモリ206上の空き領域がある一定量を下回った場合などである。

【0034】

CPU201は、キャッシュメモリ206からFPC205へのデータ転送処理を行うようにデータコントローラ500を起動し、データコントローラ500によるデータの転送処理が開始される。データコントローラ500はキャッシュメモリ206からデータのブロックを読み出し、ECCによってキャッシュメモリ206から読み出したデータに誤りがないかどうかを確認する(S3006)。データコントローラ500は書き込みデータに対してLRC符号を計算し、書き込みデータに付加されたLRC符号と、計算したLRC符号とを比較して、書き込みデータに誤りがないことを確認する(S3007)。データコントローラ500はLRC符号が付加された書き込みデータについてのパリティを付加してFPC205に転送する(S3008)。FPC205はデータを受信すると、データに付加されているパリティを用いて受信したデータに誤りがないことを確認する(S3009)。FPC205はLRC符号が付加された書き込みデータについてCRC符号を計算し、ブロック書き込み要求を生成し、ブロック書き込み要求に計算したCRC符号を付帯させる(S3010)。FPC205は書き込みデータを付随させたブロック書き込み要求をディスクドライブ300に送信する。ディスクドライブ300はブロック書き込み要求に付帯されているCRC符号を用いて書き込みデータに誤りがないことを確認し(S3011)、データを記憶する。こうしてディスクドライブ300には、書き込みデータにブロック単位でLRC符号が付加されて記憶される。

【0035】

データの読み出しの際には、ディスクアレイ装置200内では書き込みと逆の処理が行われる。FPC205はCRC符号によって、ディスクドライブ300から送信されるLRC符号が付加されたデータのブロックに誤りがないことを確認する。データコントローラ500はパリティによって、FPC205から受信するデータに誤りがないことを確認する。データコントローラ500は読み出しデータについてLRC符号を計算し、読み出しデータに付加されたLRC符号を用いて読み出しデータに誤りがないことを確認する。データコントローラ500はECCを付加してキャッシュメモリ206にLRC符号の付加された読み出しデータを書き込む。データコントローラ500はキャッシュメモリからLRC符号の付加された読み出しデータを読み出し、ECCによって読み出したデータに誤りがないことを確認する。データコントローラ500は読み出しデータに付加されたLRC符号によって読み出しデータに誤りがないことを確認する。FPC204はパリティによって、データコントローラ500から転送される読み出しデータの誤りを検査する。FPC204は読み出しデータについてCRC符号を生成し、情報処理装置100にファイバチャネルプロトコルに従って読み出しデータを送信する。情報処理装置100は受信し

たデータに対して付加されたC R C 符号を用いて、読み出しデータに誤りがないことを確認する。

【0036】

このように、ディスクアレイ装置200では、C R C 符号やパリティ、E C Cなどによって、転送されるデータに対する誤り検出を行いながら、データコントローラ500はブロック単位のデータに付加されるL R C 符号を用いたデータの誤り検出が行っている。したがって、ディスクアレイ装置200は信頼性の高いデータの記憶サービスを提供できるようになっている。

【0037】

====ブロック単位で伝送されない状況の説明====

ファイバチャネルやi S C S Iなどのプロトコルによって、複数の情報処理装置100が1つのディスクアレイ装置にアクセスすることができるようになり、1ブロック分のデータが連続して到達しないような状況が起こりうるようになってきた。このような状況を図4及び図5を用いて説明する。

【0038】

図4は、ファイバチャネルプロトコルによるフレームの大きさと、ディスクドライブ300の記憶するデータのブロックの大きさとが異なる場合を示した模式図である。S A N 4 0 0 上で伝送されるデータの単位（以下、フレームと称する）が、ディスクドライブ300上に管理される論理ブロックの大きさよりも大きく、スイッチなどにより複数のフレームが適宜混在して伝送されるような場合、S A N 4 0 0 上で伝送されるフレームには、複数の論理ブロックに対する書き込みデータが混在することがある。図4では、論理ブロックAに対する書き込みデータと、論理ブロックBに対する書き込みデータとが同一のフレーム1に混在している様子が示されている。論理ブロックAに対する書き込みデータは全体がフレーム1に含まれているが、論理ブロックBに対する書き込みデータはその一部しかがフレーム1に含まれていない。したがって、ディスクアレイ装置200はフレーム1を受信しても、データコントローラ500がキャッシュメモリ206に転送するべき書き込みデータである論理ブロックBのデータをブロック単位で受信することができない。

【0039】

また、例えば、ディスクアレイ装置200が複数のF P C 2 0 4を備えており、それら複数のF P C 2 0 4が同一のバス207を介して受信した書き込みデータをデータコントローラ500に伝送するような構成である場合にも、データコントローラ500は複数のF P C 2 0 4からの書き込みデータが混在するデータを受信することがある。

【0040】

図5は、ディスクアレイ装置200が複数のF P C 2 0 4を備え、異なる論理ブロックに対するデータが混在してデータコントローラ500に伝送される状態を示す説明図である。F P C 1 (204)がデータ「A 0」及び「A 1」を送信し、F P C 2 (204)がデータ「B 0」及びデータ「B 1」を送信すると、バス207上にはこれらのデータが混在し、図5に示すように、データコントローラ500にはデータ「A 0」「B 0」「B 1」「A 1」の順番で書き込みデータが到着することもある。

【0041】

このように、ディスクアレイ装置200が論理ブロックに対するデータを論理ブロックの一部のみしか受信できない場合や、情報処理装置100がディスクアレイ装置200に対して可変長のデータを送信するような場合など、ディスクアレイ装置200がディスクドライブ300に対する書き込みデータを、論理ブロックの連続する領域に書き込む順序に連続して受信することができないことがある。

【0042】

====本発明の転送フロー====

本発明では、ディスクアレイ装置200が論理ブロックに対する書き込みデータが論理ブロックの連続する領域に書き込む順序に連続して受信できないような場合であっても、データコントローラ500が論理ブロック単位に保証コードを付加できるようにしている

。以下にその仕組みを説明する。

【0043】

図6は、データコントローラ500がFPC204から書き込みデータを受信し、キャッシュメモリ206に書き込む処理の流れを説明するフローチャートである。

【0044】

データコントローラ500は、あらかじめキャッシュメモリ206から、ある論理ブロックのデータを読み出し、データ記憶部504に記憶しておく。

【0045】

データコントローラ500は、FPC204からバス207を介して、ディスクドライブ300のアドレスが指定された1ワードのデータを受信する(S6001)。データコントローラ500は指定されたアドレスから書き込み先となる論理ブロックアドレスを求め、書き込み先の論理ブロックと、データ記憶部504に記憶されている論理ブロックとが同一のものであるかどうかを判断する(S6002)。それらが異なる場合には(S6002: NO)、データコントローラ500はデータ記憶部504に記憶されている論理ブロックのデータについて保証コードの計算を行う(S6003)。データコントローラ500はデータ記憶部504に記憶されている論理ブロックのデータに計算した保証コードを付加して、キャッシュメモリ206に転送する(S6004)。データコントローラ500は、受信したデータの書き込み先となる論理ブロックをキャッシュメモリ206から読み出す(S6005)。データコントローラ500は、指定されたアドレスに応じてデータ記憶部504に記憶されているデータを更新する(S6006)。

【0046】

データコントローラ500が受信した書き込みデータの書き込み先となる論理ブロックと、データ記憶部504に記憶しているデータの論理ブロックとが同一である場合(S6002: YES)には、データコントローラ500はそのままデータ記憶部504に記憶されているデータを更新する(S6006)。

【0047】

なお、データコントローラ500が書き込み対象となる論理ブロックのデータをキャッシュメモリ206から読み込むために、例えば、CPU201はキャッシュメモリ206上にデータをあらかじめディスクドライブ300からステージングしておくようとする。ステージングのタイミングは、例えば、FPC204がブロック書き込み要求を受信した時点や、データコントローラ500が各ワード単位のデータを受信した時点、あるいは(S6005)の処理の直前とすることができます。

【0048】

====保証コードの計算====

データコントローラ500が論理ブロックのデータについて計算する保証コードは、排他的論理和を用いたLRC符号である。図7及び図8を用いてLRC符号の計算について説明する。本実施の形態では、図8に示すように、ブロックのデータをデータ記憶部504に記憶し、保証コードを計算する。

【0049】

図7及び図8は、データコントローラ500の備えるデータ記憶部504に記憶される1ブロック分のデータを示す模式図である。図7では、データ記憶部504には1ブロック512バイトのデータが2バイトずつアドレス順に並び、末尾には、論理ブロックアドレスの下位8ビット(LBA0)と論理ユニット番号の下位8ビット(LUN0)とが付加されている。図8では、データ記憶部504には1ブロック512バイトのデータが4バイトずつアドレス順に並び、末尾には、論理ユニット番号の下位16ビット分(LUN0及びLUN1)と、論理ブロックアドレスの下位16ビット分(LBA0及びLBA1)とが付加されている。本実施の形態では、図8に示すデータがキャッシュメモリ206及びディスクドライブ300に記憶されるようにしている。

【0050】

LRC符号の計算を始める前に、データコントローラ500は計算するLRC符号の初

期値として「0」を設定する。よって、図7ではLRC0及び1の値が「0」となり、図8ではLRC0乃至3の各バイトが「0」となる。

【0051】

データコントローラ500は、図7に示す演算方向801に沿って、データ記憶部504に記憶する1ブロック分のデータについて、ディスクドライブ300のアドレス順に2バイトずつのデータとLRC符号（LRC0及び1）とを用いて排他的論理和を行う。図7では、16ビットのLRC符号（LRC0及び1）が求められ、LBA0及びLUN0とともに、32ビットの保証コードとしてキャッシングメモリ206に書き込まれる。

【0052】

一方、図8に示す演算方向901に沿う演算では、データコントローラ500はLRC符号に対して、上記4バイト単位のデータを排他的論理和演算していく。データコントローラ500はその演算の結果をLRC符号としてLRC0乃至3に記憶する。ここで、データコントローラ500は、n番目の4バイト単位のデータを（8×n）ビット左に巡回シフトしてから排他的論理和演算を行う。よって、図8に示すバイト単位のデータについてみると、例えば、LRC1の値は、演算方向901にしたがって、「A2xA5x...x A511xLBA0」として求められる。ここで「x」は排他的論理和演算を表す。

【0053】

図7の演算方向801に従って計算するLRC符号は、容易に計算できる反面、データの順序についての保証を行うことができない。例えば、連続するデータが間違った送信順序でデータコントローラ500に到達した場合、本来2番目に到達するべきデータが1番目に送信されてきたとしても、排他的論理和演算の演算結果は、本来の順番のデータで受信した場合の演算結果と同じになってしまう。

【0054】

そこで、図8に示す演算方向901のように、データコントローラ500はデータ記憶部504に記憶されているデータに対して巡回シフトしながら排他的論理和演算をすることによって、FPC204から受信したデータが指定されたアドレスの順序に正しく並んでいない場合に排他的論理和演算の結果が正しい保証コードになってしまふ不具合を防ぐようにしている。

【0055】

なお、排他的論理和演算に際して各4バイトのデータを巡回シフトする単位は8ビットに限らず、1ビット、2ビット、4ビットなど任意のビット数をシフトするようにしてもよい。

【0056】

====その他の実施の形態====

本実施の形態において、データコントローラ500がディスクドライブ300に対する書き込みデータのブロック単位に保証コードを付加していたが、これはFPC204で実現することもできる。

【0057】

FPC204がデータ記憶部を備え、ディスクドライブ300から1ブロック分のデータを読み出す。FPC204は情報処理装置100から受信したブロック書き込み要求に応じて、FPC204の備えるデータ記憶部に記憶した1ブロックのデータを更新し、保証コードを計算する。データコントローラ500は、保証コードの付加されたブロックデータをキャッシングメモリ206に転送する。データ記憶部はFPC204に備える形態としてもよいし、FPC204がアクセスできるメモリ202などの記憶装置を利用する形態としてもよい。

【0058】

また、データのブロックに対する保証コードの付加は、CPU201が行ってもよい。CPU201はディスクドライブ300に記憶されている論理ブロックのデータをメモリ202に読み出す。FPC204がブロック書き込み要求を受信すると、データコントローラ500はメモリ202に書き込みデータを転送し、メモリ202に記憶されている論

理ブロックのデータを更新する。CPU201はメモリ202に記憶されている論理ブロックのデータについての保証コードの計算を行い、メモリ202に記憶されている論理ブロックのデータの末尾に保証コードを書き込む。データコントローラ500はメモリ202に記憶されている、論理ブロックのデータと保証コードとをキャッシュメモリ206に転送する。このようにして、CPU201は、データコントローラ500と同様に、ブロック単位に保証コードの計算を行うことができる。

【0059】

本実施の形態において、保証コードはLRC符号を用いたが、保証コードにはLRC符号以外にも、CRC符号やVRC符号、ハミング符号、チェックサム、パリティなどを用いてもよい。

【0060】

本実施の形態において、ディスクアレイ装置200はFPC204が受信したディスクドライブ300に対する書き込みデータを、キャッシュメモリ206に格納し、キャッシュメモリ206からディスクドライブ300に転送する形態としたが、FPC204が受信したデータを、FPC205を介して直接ディスクドライブ300に書き込む形態としてもよい。

【0061】

====第2の実施の形態====

次に本発明の別の実施の形態のひとつであるディスクアレイ装置について説明する。図9が本第2の実施の形態にかかる情報処理システムの構成を示すブロック図である。

図9に示す情報処理システムは、上述した第1の実施の形態の情報処理システムと同様の構成となっている。ディスクアレイ装置200はディスクドライブ300の記憶領域を情報処理装置100に提供し、情報処理装置100は適宜ディスクアレイ装置200に対してブロックアクセス要求を送信し、ディスクドライブ300の記憶領域へのアクセスを行う。本実施の形態では、ディスクドライブ300は記憶デバイス310として、ディスクアレイ装置200と接続されている。記憶デバイス310とディスクアレイ装置200とはSCSIやファイバチャネルのプロトコルに従った通信経路で接続されている。記憶デバイス310とディスクアレイ装置200とは直接接続する形態としてもよいし、ネットワーク接続の形態としてもよい。また、記憶デバイス310はディスクアレイ装置200に一体的に組み込む形態としてもよい。

【0062】

ディスクアレイ装置200は、図9に示すように、チャネル制御部210、共有メモリ220、スイッチ230、キャッシュメモリ206、ディスク制御部240、管理端末250などを備えている。

【0063】

チャネル制御部210はSAN400に接続され、情報処理装置100との間で通信を行う。チャネル制御部210は通信インターフェースを備え、例えば、LANやSCSI、iSCSI、ESCON（登録商標）、FICON（登録商標）、ACONARC（登録商標）、FIBARC（登録商標）などのプロトコルにしたがって、ディスクドライブ300に対するデータ入出力を指示するコマンド（以下、ディスク制御コマンドと称する）をディスク制御部240に伝達する。ディスク制御部240はチャネル制御部210からディスク制御コマンドを受け取ると、ディスクドライブ300に対するデータ入出力に関する制御を行う。チャネル制御部210からディスク制御部240に対するディスク制御コマンドは、共有メモリ220を介して伝達される。共有メモリ220はチャネル制御部210やディスク制御部240、管理端末250などによって共有されるメモリである。

【0064】

管理端末250はディスクアレイ装置200を保守管理するためのコンピュータである。管理端末250はディスクアレイ装置200に内臓された形態とすることもできるし、独立したコンピュータ端末としてディスクアレイ装置200に接続する形態とすることもできる。チャネル制御部210やディスク制御部240において実行されるソフトウェア

やデータの更新、設定情報の変更などは管理端末250からの指示によって行われる。

【0065】

====チャネル制御部====

図10は、チャネル制御部210の構成を示すブロック図である。チャネル制御部210は一体的にユニット化されたボードで構成されている。チャネル制御部210はCPU211、メモリ212、NVRAM213、通信インターフェース部214、データコントローラ500などを備え、これらが一体的にユニット化された回路基板上に形成されて構成されている。

【0066】

通信インターフェース部214は、情報処理装置100との間で通信を行うためのインターフェースを備えている。通信コネクタ215は情報処理装置100と通信を行うためのコネクタである。本実施の形態におけるチャネル制御部210の場合、通信コネクタ215は、SAN400に接続可能なコネクタであり、例えば、ファイバチャネルに対応している。チャネル制御部210が情報処理装置100からファイル名を指定したデータ入出力要求を受け付けるのであれば、通信コネクタは例えばイーサネット（登録商標）に対応し、チャネル制御部210はLAN経由でデータ入出力要求を受け付ける。

【0067】

接続コネクタ216はチャネル制御部210がディスクアレイ装置200と接続するためのコネクタである。接続コネクタ216がディスクアレイ装置200側のコネクタと嵌合することにより、チャネル制御部210のボードはディスクアレイ装置200と電気的に接続される。チャネル制御部210は接続コネクタ216を介してスイッチ230に接続され、ディスクアレイ装置210内の共有メモリ220やキャッシュメモリ206、ディスク制御部240などにアクセスが可能となっている。

【0068】

CPU211はチャネル制御部210全体の制御を司る。CPU211はNVRAM213に記憶されたアプリケーションプログラムをメモリ212に読み出して実行することによって、各種の機能が実現される。NVRAM213は各種プログラムや設定データなどを格納する不揮発性のメモリである。NVRAM213に記憶されるアプリケーションプログラムや設定データなどの内容は管理端末250からの指示により書き換えを行うことができる。

【0069】

上述した第1の実施の形態で説明したデータコントローラ500は、このチャネル制御部210内においてデータの転送を制御するようにしている。データコントローラ500の構成は上述した図2に示す部ブロック構成と同様となっている。

【0070】

チャネル制御部210が情報処理装置100からブロック書き込み要求を受信すると、CPU211の制御によって、通信インターフェース部213が受信したデータをワード単位にデータコントローラ500に伝送する。データコントローラ500は通信インターフェース部213から受け取ったデータを、上に説明した図6に示す処理の流れと同様に、キャッシュメモリ206に書き込んでいく。キャッシュメモリ206に書き込まれたデータのブロックには論理ブロック単位で保証コードが付加されていることになる。

【0071】

キャッシュメモリ206に書き込まれた書き込みデータは、ディスク制御部240によってディスクドライブ300に書き込まれる。これにより、ディスクドライブ300に書き込まれるデータには、論理ブロック単位で保証コードが付加されている。

【0072】

ディスクアレイ装置200は、情報処理装置100からデータを読み出すためのブロックアクセス要求を受信すると、チャネル制御部210がその要求に応じてディスク制御部240に対してデータを読み出すように指示するコマンドを生成し、共有メモリ220に書き込む。ディスク制御部240は共有メモリ220から読み出したコマンドに応じて、

ディスクドライブ300からデータをブロック単位に読み出し、キャッシュメモリ206に書き込む。このとき読み出したブロック単位のデータには、末尾に保証コードが付加されている状態でキャッシュメモリ206に書き込まれることになる。チャネル制御部210はキャッシュメモリ206からブロック単位のデータを読み出し、読み出したデータに付加されている保証コードを用いてデータの完全性を検査する。チャネル制御部210は保証コードを除いたブロック単位のデータを情報処理装置100に送信する。このようにして、ディスクアレイ装置200は、データの誤りを適宜検査しながらデータの読み書きを行うことによって、信頼性の高いデータの記憶サービスを提供することができる。

【0073】

====ディスク制御部====

書き込みデータに対して保証コードを付加する処理は、ディスク制御部240で行うようになることもできる。図11は、ディスク制御部240の構成を示すブロック図である。ディスク制御部240は、一体的にユニット化されたボードとして形成されている。ディスク制御部240のボードは、インターフェース部241、CPU242、メモリ243、NVRAM244、接続コネクタ245、データコントローラ500を備え、これらが一体的にユニット化された回路基板上に形成されている。

【0074】

ディスクインターフェース部241は、ディスクドライブ300との間で通信を行うための通信インターフェースを備えている。コネクタ242はディスクドライブ300と通信を行うためのコネクタである。コネクタ242は、例えば、ディスクドライブ300がSCSIのハードディスクである場合、SCSIのコネクタであり、ディスクドライブ300がファイバチャネルの規約にしたがうハードディスクであれば、ファイバチャネルに接続するコネクタとなる。

【0075】

CPU242は、ディスク制御部240全体の制御を司ると共に、チャネル制御部210やディスクドライブ300、管理端末250との間の通信を行う。メモリ243やNVRAM244に格納された各種プログラムを実行することにより本実施の形態に係るディスク制御部240の機能が実現される。ディスク制御部240により実現される機能としては、ディスクドライブ300の制御やRAID制御、ディスクドライブ300に記憶されたデータの複製管理やバックアップ制御、リモートコピー制御等である。

【0076】

NVRAM244はCPU242の制御を司るプログラムを格納する不揮発性メモリである。NVRAM244に記憶されるプログラムの内容は、管理端末250からの指示により書き込みや書き換えを行うことができる。

【0077】

接続コネクタ245がディスクアレイ装置200側のコネクタと嵌合することにより、ディスク制御部240のボードはディスクアレイ装置200と電気的に接続される。ディスク制御部240はボード接続コネクタ216を介してスイッチ230に接続し、ディスクアレイ装置210内の共有メモリ220やキャッシュメモリ206、チャネル制御部210などにアクセスが可能になっている。

【0078】

データコントローラ500は、これまでに説明したデータコントローラ500と同様のものである。ただし、ディスク制御部240の備えるデータコントローラ500は、キャッシュメモリ206から読み出すデータを受信データとし、保証コードを付加したブロックのデータはディスクドライブ300に書き込まれることとなる。

【0079】

チャネル制御部210は保証コードを付加せずに、受信したデータをキャッシュメモリ206に書き込む。ディスク制御部240はキャッシュメモリ206から読み出すデータを受信データとして、図6に示す処理の流れにしたがって、保証コードを計算し、キャッシュメモリ206ではなくディスクドライブ300にデータを書き込む。

【0080】

また、チャネル制御部210がディスクドライブ300に対する書き込みデータ受信し、キャッシュメモリ206上のデータを更新し、ディスク制御部240が保証コードを付加するようにしてもよい。具体的には、チャネル制御部210は情報処理装置からディスクドライブ300に対する書き込みデータを受信すると、ディスク制御部240がキャッシュメモリ206上にデータをステージングする。チャネル制御部210は、受信したデータを隨時キャッシュメモリ206に書き込んで、ステージングしたデータを更新する。ディスク制御部240はキャッシュメモリ206からステージングしたデータを読み出し、ブロック単位に保証コードを計算してディスクドライブ300に書き込む。

【0081】

以上説明したように、本実施の形態にかかるディスクアレイ装置200によれば、ディスクアレイ装置200に連続して送信されてくる書き込みデータが論理ブロックの連続した領域に書き込まれる順序でディスクアレイ装置200に到達しなかったとしても、ディスクアレイ装置200はディスクドライブ300に記憶されるデータにブロック単位で保証コードを付加することができる。したがって、信頼性の高いデータの記憶サービスを提供することができる。

【0082】

以上本実施の形態について説明したが、上記実施の形態は本発明の理解を容易にするためのものであり、本発明を限定して解釈するためのものではない。本発明はその趣旨を逸脱することなく変更、改良され得ると共に、本発明にはその等価物も含まれる。

【図面の簡単な説明】

【0083】

【図1】本発明の一実施の形態による、データ転送回路を備えるディスクアレイ装置を含む情報処理システムの構成を示すブロック図である。

【図2】本発明の一実施の形態による、ディスクアレイ装置200の備えるデータコントローラ500のブロック図である。

【図3】本発明の一実施の形態による、情報処理装置100からブロック書き込み要求を受信し、ディスクドライブ300に書き込むまでの流れを示す説明図である。

【図4】本発明の一実施の形態による、ファイバチャネルプロトコルによるフレームの大きさと、ディスクドライブ300の記憶するデータのブロックの大きさとが異なる場合を示した模式図である。

【図5】本発明の一実施の形態による、ディスクアレイ装置200が複数のFPC204を備え、異なる論理ブロックに対するデータが混在してデータコントローラ500に伝送される状態を示す説明図である。

【図6】本発明の一実施の形態による、データコントローラ500がFPC204から書き込みデータを受信し、キャッシュメモリ206に書き込む処理の流れを説明するフローチャートである。

【図7】本発明の一実施の形態による、データコントローラ500の備えるデータ記憶部504に記憶される1ブロック分のデータを示す模式図である。

【図8】本発明の一実施の形態による、データコントローラ500の備えるデータ記憶部504に記憶される1ブロック分のデータを示す模式図である。

【図9】本発明の一実施の形態による、情報処理システムの構成を示すブロック図である。

【図10】本発明の一実施の形態による、チャネル制御部210の構成を示すブロック図である。

【図11】本発明の一実施の形態による、ディスク制御部240の構成を示すブロック図である。

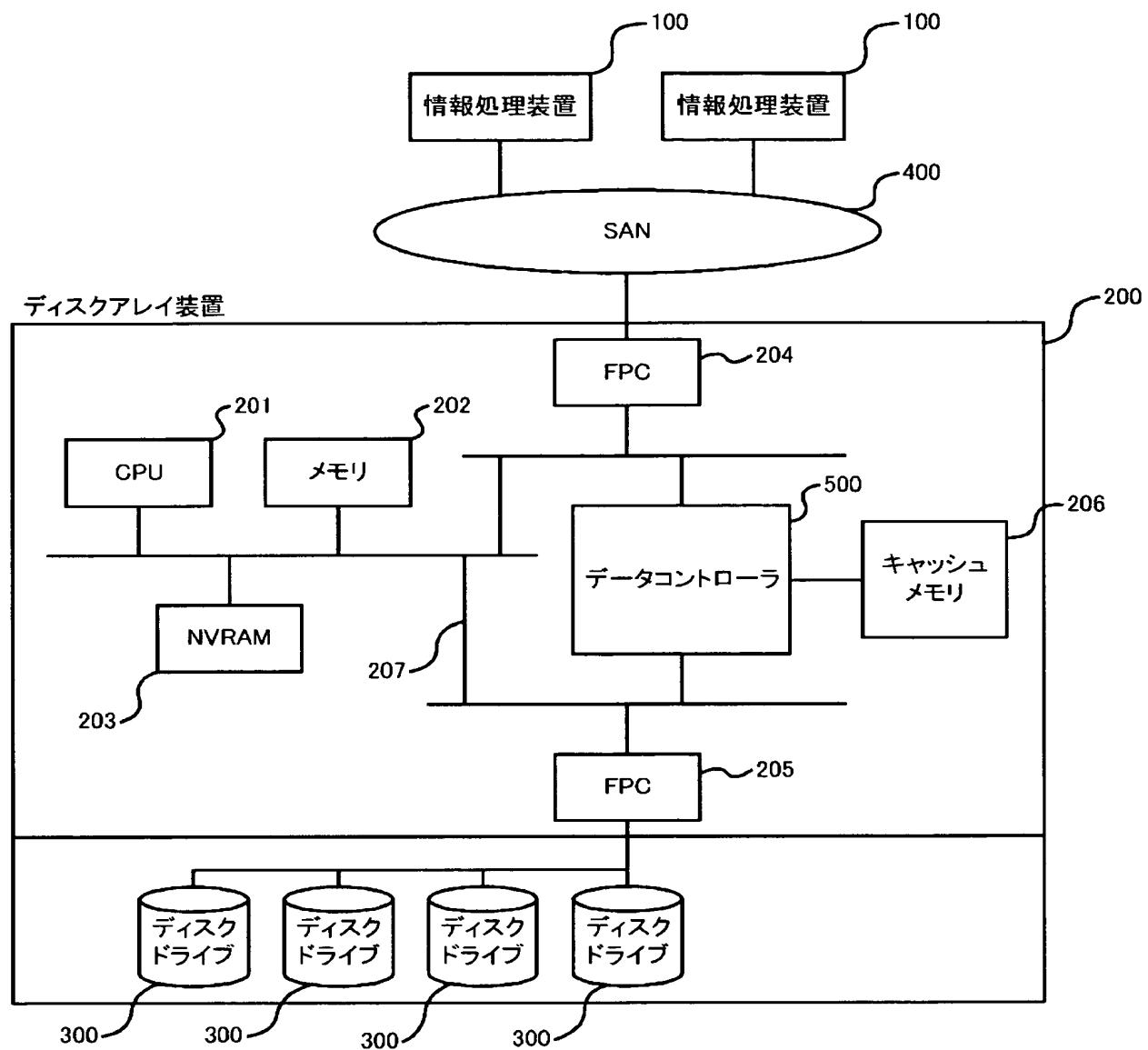
【符号の説明】

【0084】

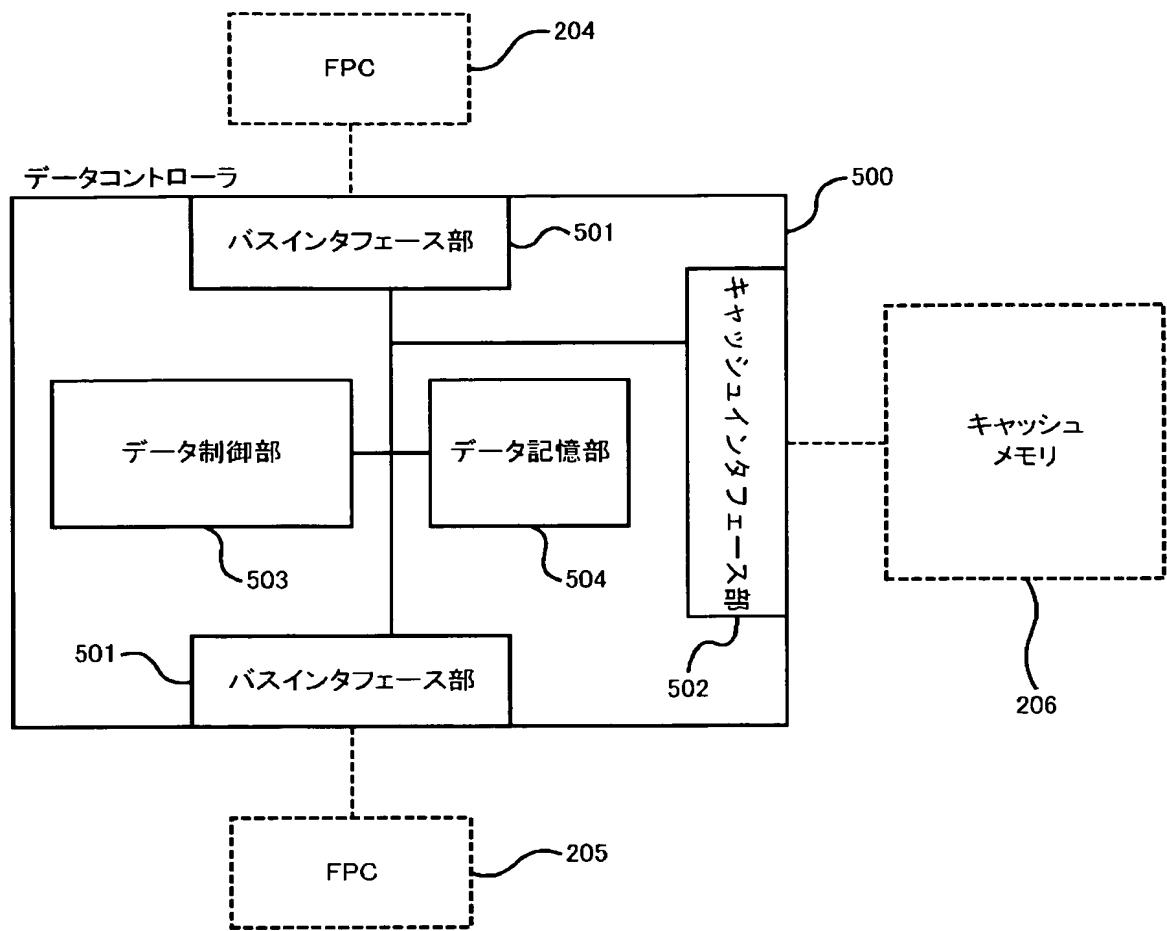
100 情報処理装置

200 ディスクアレイ装置
201 C P U
202 メモリ
203 N V R A M
204 F P C (ホストインターフェース)
205 F P C (ディスクインターフェース)
206 キャッシュメモリ
207 バス
210 チャネル制御部
220 共有メモリ
230 スイッチ
240 ディスク制御部
250 管理端末
300 ディスクドライブ
310 記憶デバイス
400 S A N
500 データコントローラ
501 バスインターフェース部
502 キャッシュインターフェース部
503 データ制御部
504 データ記憶部

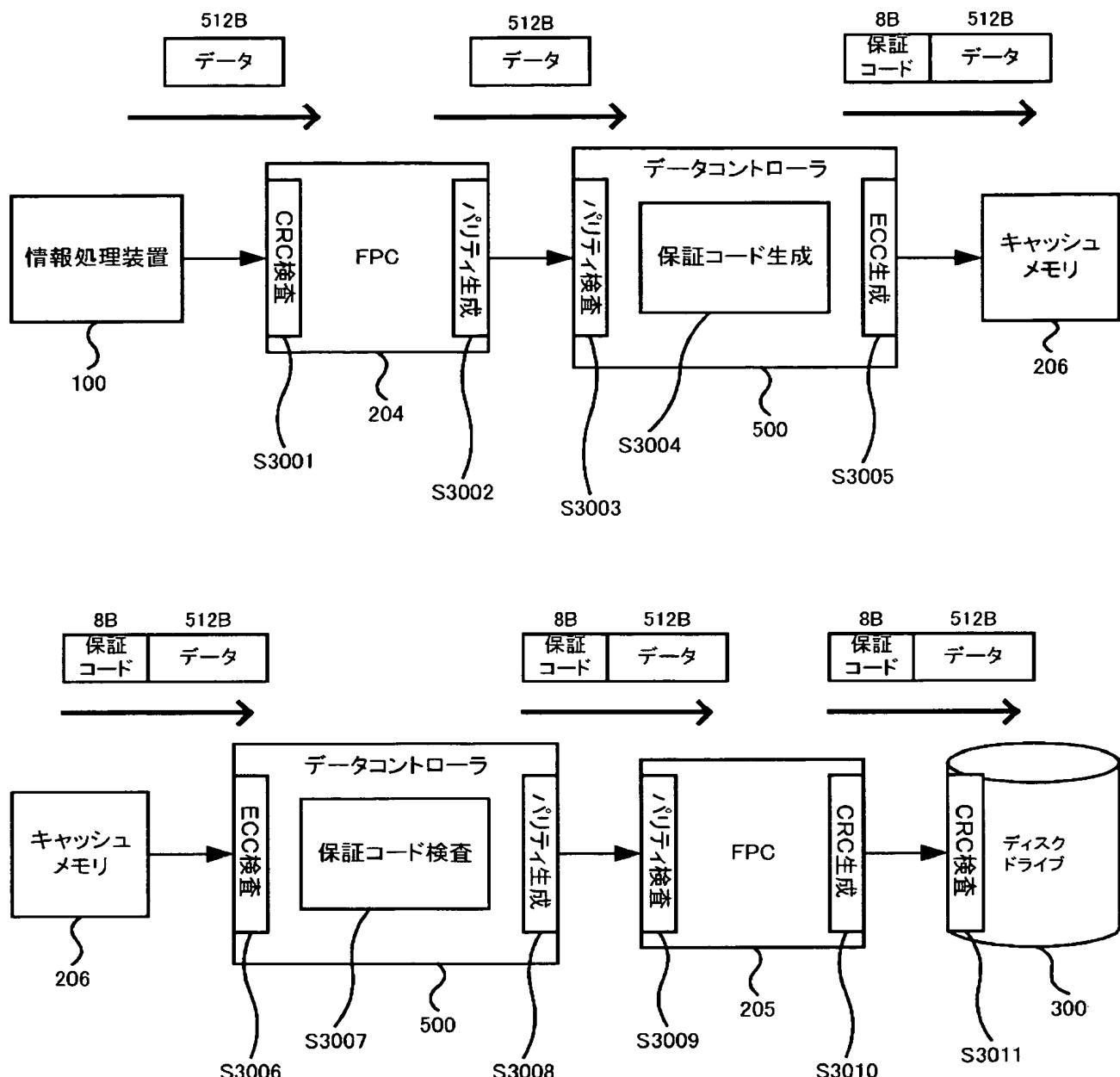
【書類名】図面
【図1】



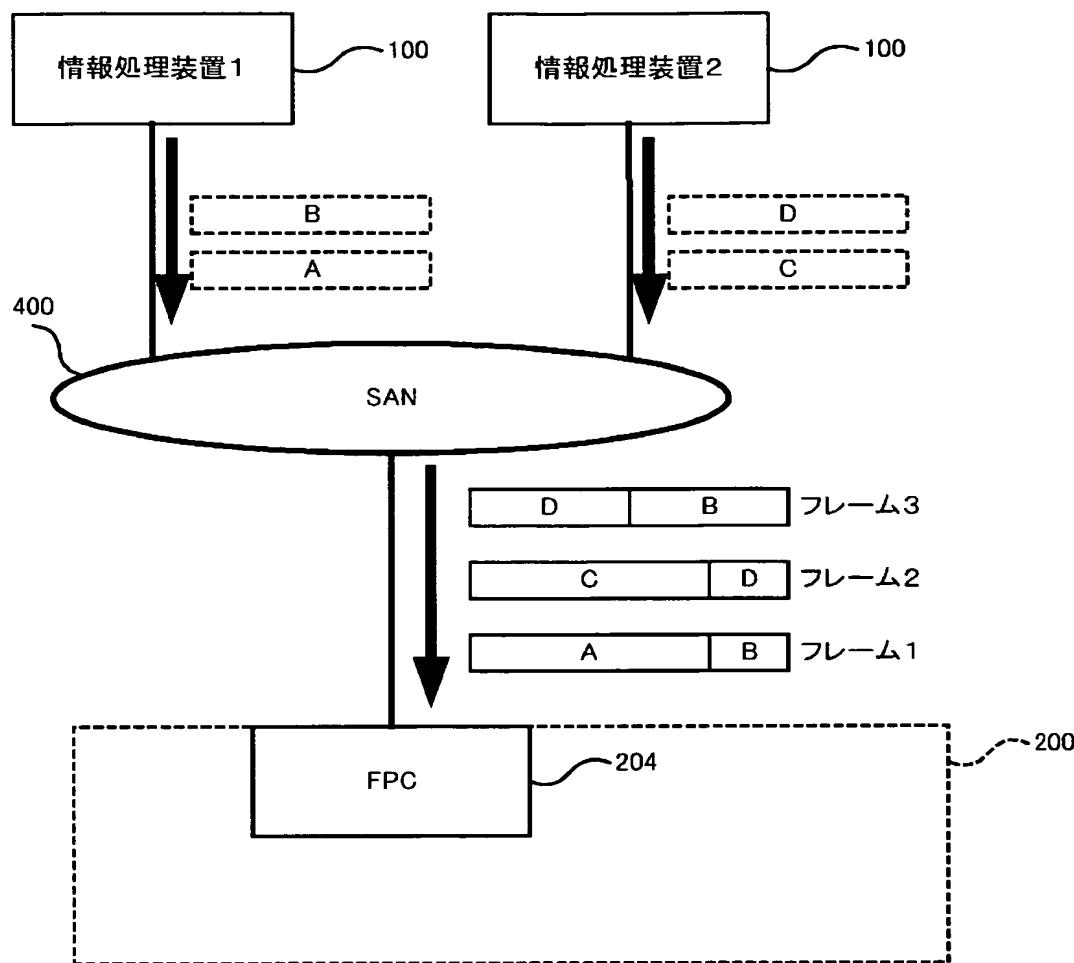
【図2】



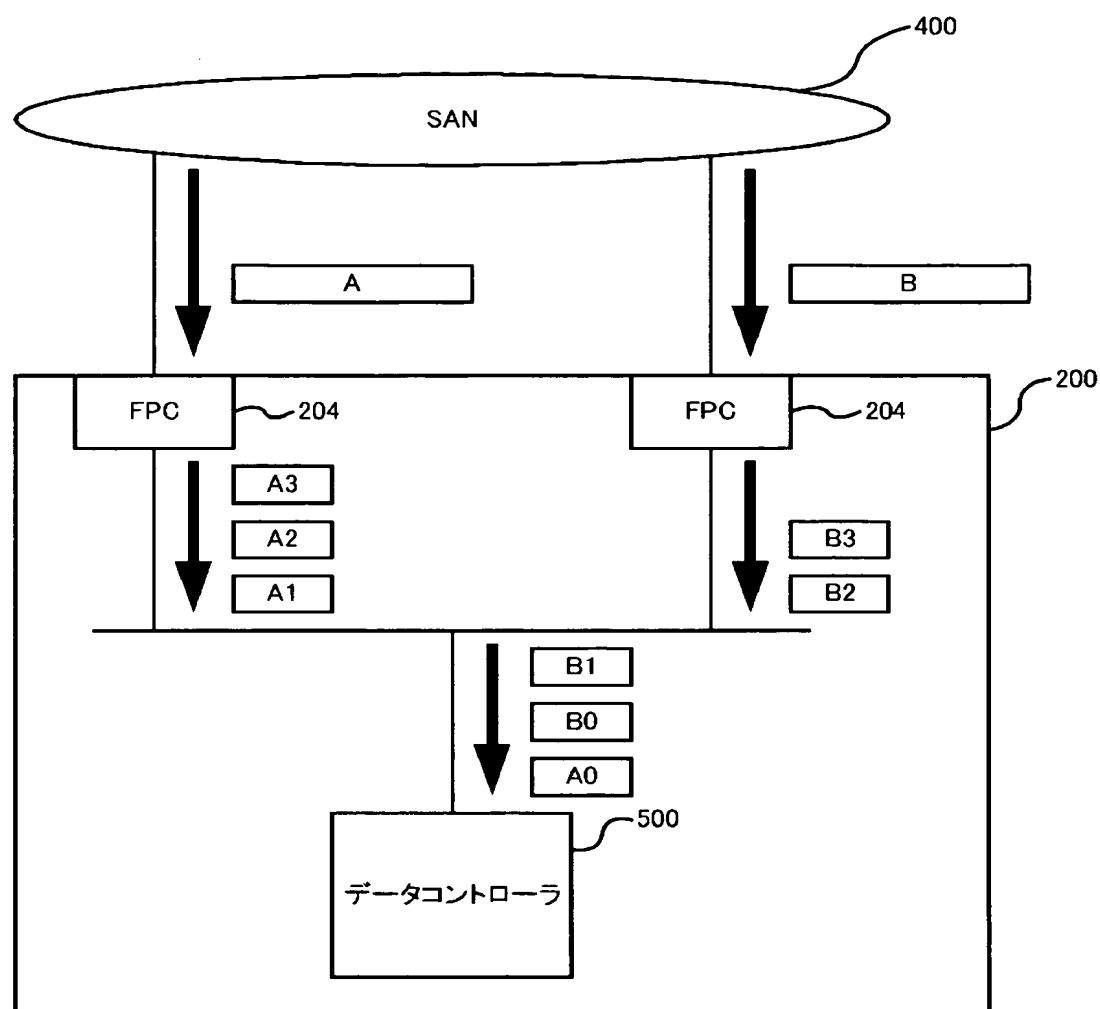
【図3】



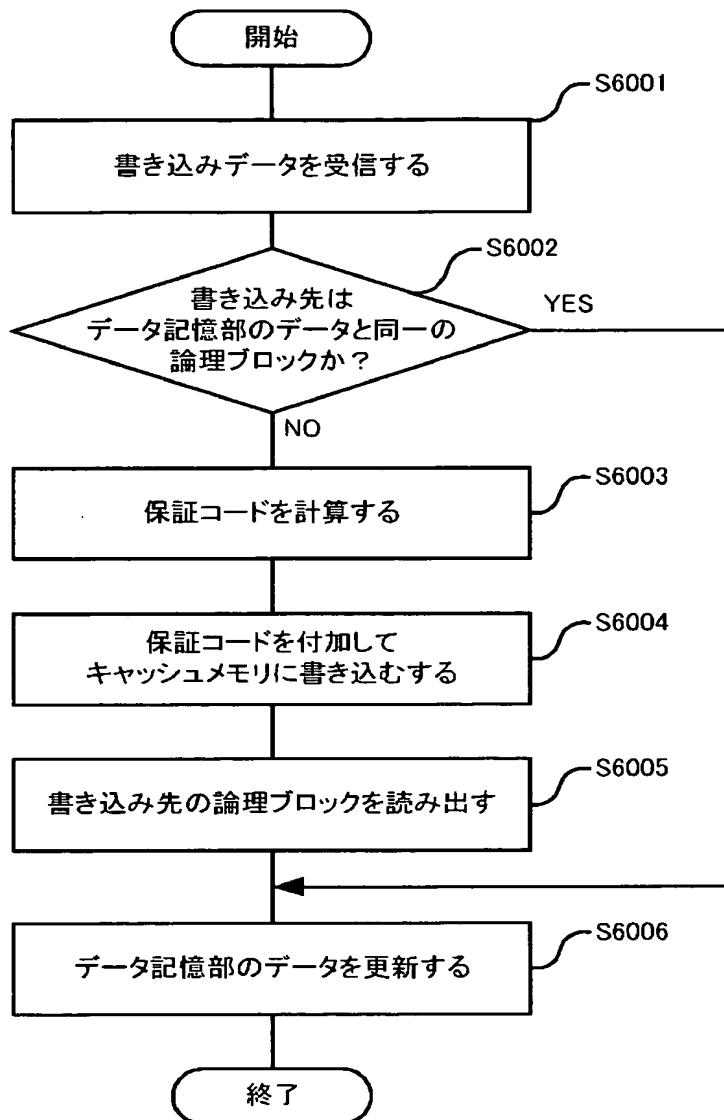
【図4】



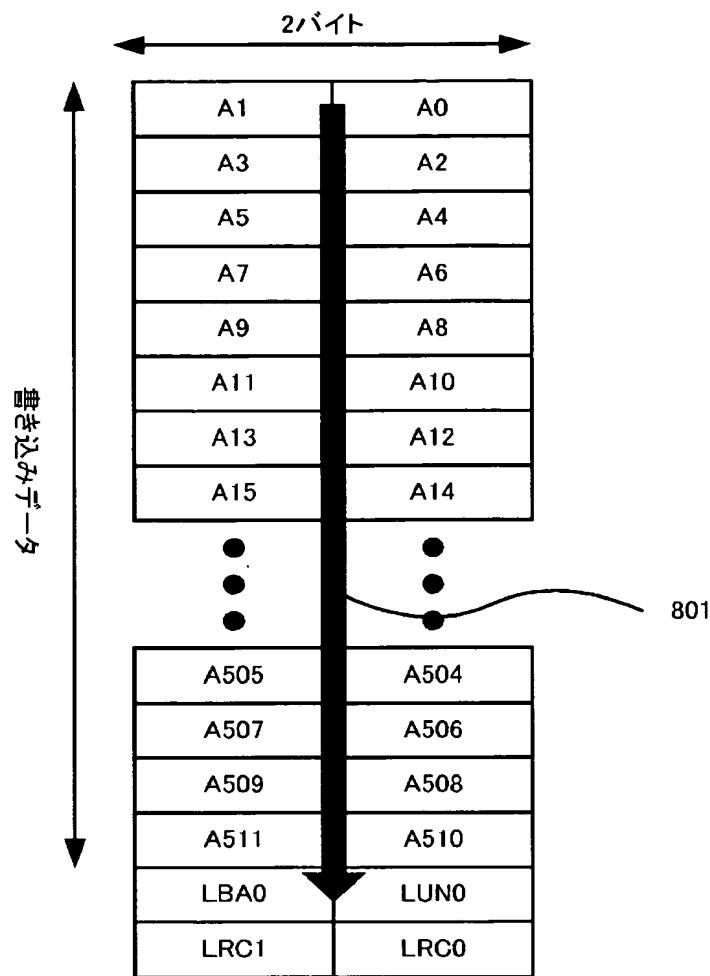
【図5】



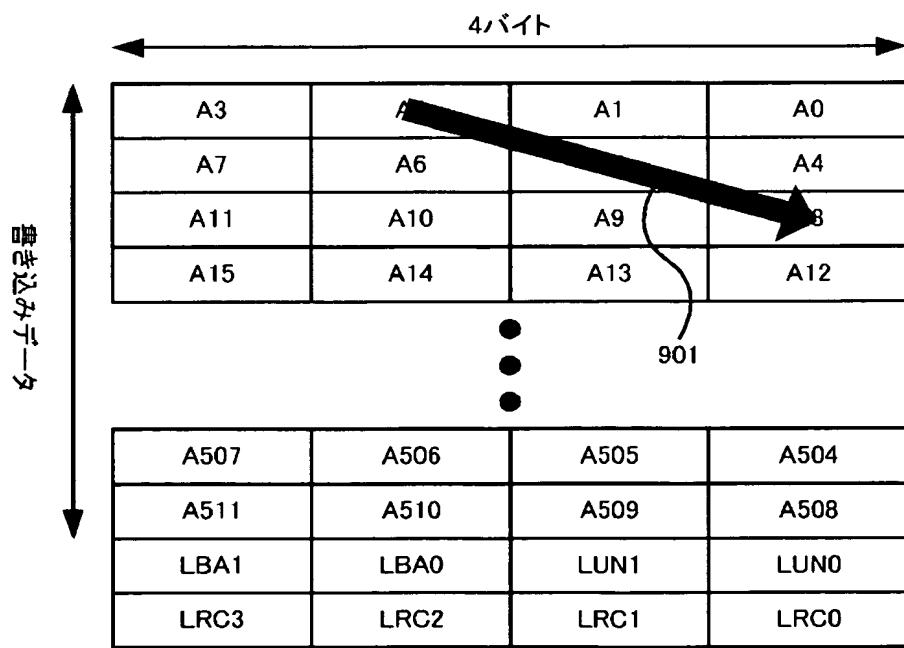
【図6】



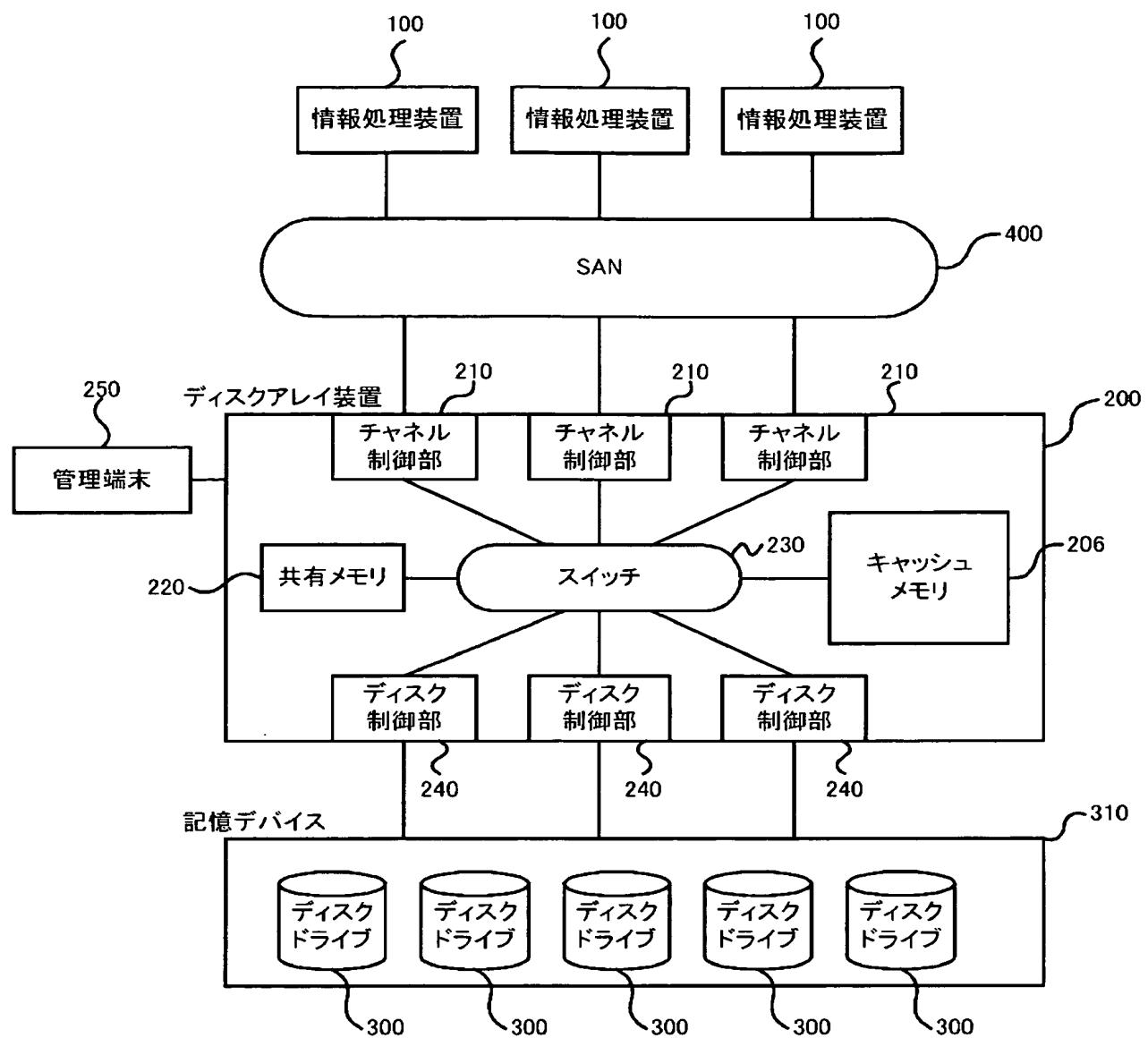
【図7】



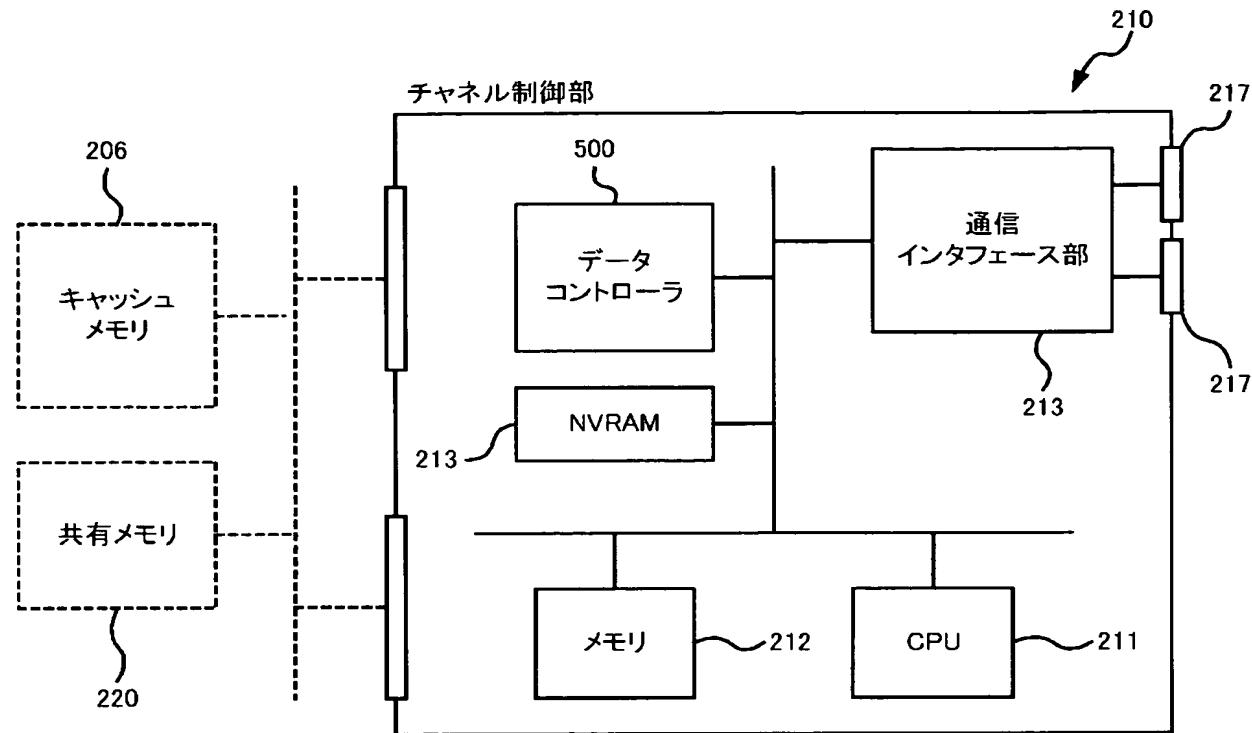
【図8】



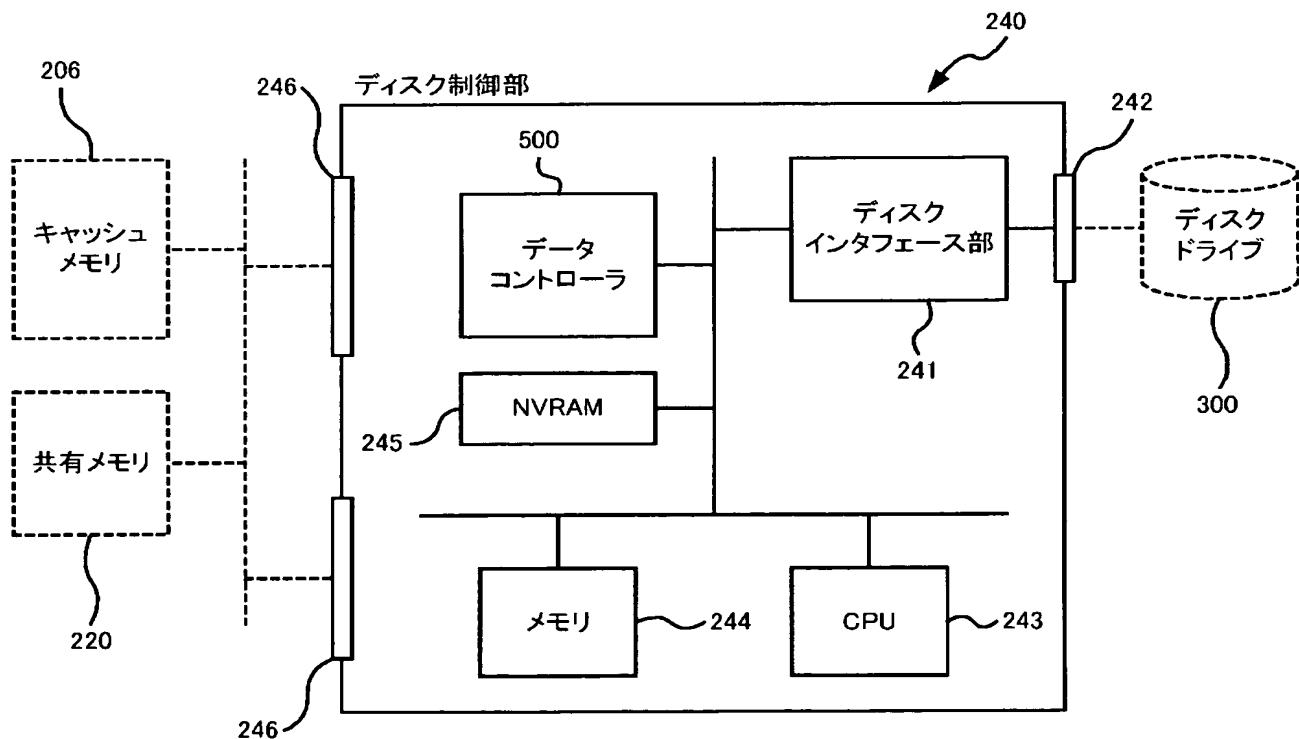
【図9】



【図10】



【図11】



【書類名】要約書

【要約】

【解決手段】 記憶装置に対する書き込みデータを受信するデータ受信部と、データ受信部が受信する書き込みデータを記憶装置に転送するデータ制御部と、記憶装置の記憶領域上に記憶される一連のデータを記憶するデータ記憶部とを備えるデータ転送装置の制御方法である。データ制御部は記憶装置から一連のデータをブロック単位で読み出してデータ記憶部に記憶する。データ制御部は書き込みデータの書き込み先となるブロックとデータ記憶部に記憶されているブロックとが同じ場合は、データ記憶部に記憶されているデータを書き込みデータにより更新し、異なる場合は、データ記憶部に記憶されている一連のデータに基づいて保証コードを生成し、その一連のデータに保証コードを附加して記憶装置に転送する。データ制御部は書き込みデータの書き込み先となるブロックの一連のデータを読み出してデータ記憶部に記憶し、書き込みデータにより更新する。

【選択図】 図6

特願 2003-314010

出願人履歴情報

識別番号 [000005108]

1. 変更年月日 1990年 8月31日

[変更理由] 新規登録

住所 東京都千代田区神田駿河台4丁目6番地
氏名 株式会社日立製作所